

Démélange parcimonieux exact dans une approche supervisée en imagerie hyperspectrale

Ramzi BEN MHENNI¹, Sébastien BOURGUIGNON¹, Jordan NININ², Frédéric SCHMIDT³

¹École Centrale de Nantes, LS2N, 1 rue de la Noë, F-44321 Nantes, France

²ENSTA-Bretagne, Lab-STICC, 2 rue François Verny, F-29806 Brest cedex 9, France

³Université Paris-Sud, GEOPS, 1 rue du Belvédère, F-91405 Orsay, France

ramzi.ben-mehnni@ec-nantes.fr ; sebastien.bourguignon@ec-nantes.fr
jordan.ninin@ensta-bretagne.fr ; frederic.schmidt@u-psud.fr

Résumé – Nous proposons une approche d’optimisation en norme ℓ_0 pour le démelange spectral linéaire sous contrainte de parcimonie. Nous montrons que la résolution exacte est faisable, pour des problèmes de complexité limitée mais réaliste, par l’intermédiaire de la programmation mixte en nombres entiers. Des simulations numériques montrent que la prise en compte de cette parcimonie exacte permet d’améliorer la qualité de détection des abondances spectrales par rapport aux approches existantes.

Abstract – We propose an ℓ_0 -norm based optimisation approach to the sparse linear spectral unmixing problem. We show that exact solutions can be obtained through mixed-integer programming techniques, for problems with limited, yet realistic, complexity. Numerical simulations show that solving exactly such sparsity-constrained problem improves the detection performance of spectral abundances over existing methods.

1 Introduction

Un problème classique en imagerie hyperspectrale est celui de démelange spectral, où l’on cherche à décomposer chaque spectre en un mélange linéaire de spectres élémentaires *purs* et à estimer les proportions (abondances) associées. Dans une approche supervisée, on ne s’intéresse qu’à l’estimation des abondances, les spectres de référence étant supposés connus, soit par des mesures spectroscopiques en laboratoire des différents minéraux susceptibles d’être présents, soit *via* leur estimation à partir des mêmes données hyperspectrales.

Toutes les signatures n’étant pas systématiquement présentes dans chaque spectre observé (en particulier dans le premier cas, où le *dictionnaire* de spectres de référence peut contenir un nombre arbitrairement élevé de composants), il peut sembler naturel de rechercher des solutions parcimonieuses, *i.e.*, pour lesquelles le nombre de coefficients non nuls (la « norme » ℓ_0) est faible [1]. La résolution d’un problème sous contraintes de positivité produit naturellement des estimateurs comportant un certain nombre de zéros. En revanche, si l’introduction explicite de contraintes de parcimonie a fait l’objet de différents travaux [2, 3], les algorithmes classiques basés sur la norme ℓ_1 ou les approches gloutonnes n’apportent qu’une réponse partielle à ce problème, car ils sont inappropriés à la nature des variables recherchées, dont la somme vaut au plus 1.

Des travaux récents [4] ont montré, sur des problèmes de déconvolution impulsionnelle, que la résolution *exacte* du problème d’optimisation parcimonieuse en norme ℓ_0 , essentiellement combinatoire (donc *a priori* très coûteux), est envisa-

geable sur des problèmes de taille raisonnable, au moyen de méthodes appropriées reposant sur des algorithmes *Branch-and-Bound*. Cela peut être le cas en démelange spectral, où le nombre de longueurs d’onde mesurées reste souvent limité à quelques centaines et le nombre de composants recherchées dans le spectre de chaque pixel ne dépasse pas la dizaine.

Dans cet article, nous étudions d’une part la faisabilité de l’approche parcimonieuse exacte en norme ℓ_0 sur ce problème. D’autre part, nous évaluons l’apport de cette approche relativement aux méthodes de résolution approchées en termes de qualité de solution.

2 Démélange linéaire parcimonieux

L’objectif dans un problème de démelange supervisé est d’estimer les abondances de chacun des composants spectraux purs d’un spectre observé $\mathbf{y} \in \mathbb{R}^L$, acquis dans L bandes spectrales, résultant du mélange de spectres élémentaires. Si on considère que le modèle de mélange est linéaire [5], alors, on peut écrire :

$$\mathbf{y} = \sum_{n=1}^N a_n \mathbf{s}_n + \boldsymbol{\epsilon} = \mathbf{S}\mathbf{a} + \boldsymbol{\epsilon},$$

où $\mathbf{s}_n \in \mathbb{R}^L$ représente le $n^{\text{ème}}$ spectre pur, $a_n \in \mathbb{R}$ est l’abondance associée, $\mathbf{S} = [\mathbf{s}_1, \dots, \mathbf{s}_N]$, $\mathbf{a} = [a_1, \dots, a_N]$ et $\boldsymbol{\epsilon}$ représente le bruit et l’erreur de modélisation.

Pour traduire des considérations physiques (les abondances représentent des pourcentages), des contraintes de positivité et de somme à un sont en général ajoutées aux inconnues a_n .

Le problème d'estimation au sens des moindres carrés incorporant ces contraintes, souvent dénommé FCLS pour *Fully-Constrained Least-Squares* [1], s'écrit alors :

$$\min_{\mathbf{a}} \|\mathbf{y} - \mathbf{S}\mathbf{a}\|^2 \text{ t. q. } \mathbf{a} \geq 0, \sum_n a_n = 1. \quad (1)$$

Dans de nombreuses observations, les unités géologiques ne sont souvent pas composées d'un assemblage de plus de quelques minéraux majeurs. La formulation (1), par la contrainte de positivité, produit des solutions pour lesquelles certains coefficients sont nuls. Une façon possible d'imposer de manière plus explicite cette propriété de parcimonie (on n'utilise qu'un faible nombre de spectres élémentaires pour représenter le mélange) peut être décrite par la minimisation du nombre de coefficients non nuls de \mathbf{a} ($\|\mathbf{a}\|_0 = \text{Card}\{q|a_q \neq 0\}$), sous contrainte d'une borne sur l'erreur d'approximation, ou bien en imposant une borne sur le nombre d'éléments non nuls de la solution :

$$\min_{\mathbf{a} \geq 0} \|\mathbf{y} - \mathbf{S}\mathbf{a}\|^2 \text{ t. q. } \|\mathbf{a}\|_0 \leq K_{\max}, \sum_{n=1}^N a_n = 1 \quad (2)$$

Dans la suite, nous considérons cette seconde formulation, qui se prête numériquement mieux à la résolution par les approches exactes introduites en Section 3 [4].

Limites des approches classiques

La norme ℓ_0 , discrète, rend le problème de démixage (2) essentiellement combinatoire, le nombre de combinaisons augmentant exponentiellement avec le nombre de variables. Afin de limiter le coût de calcul, les approches classiques d'estimation parcimonieuse reposent soit sur la relaxation (convexe) de la norme ℓ_0 en la norme ℓ_1 , soit sur la mise en œuvre d'algorithmes gloutons. Cependant, ici, ni l'une ni l'autre ne sont appropriées. En effet, les coefficients a_n représentant des proportions, toute formulation reposant sur la norme ℓ_1 pour imposer un certain niveau de parcimonie est intrinsèquement inadaptée, puisque la norme ℓ_1 des abondances vaut 1. Notons que le problème (1) est en soi un problème d'optimisation en norme ℓ_1 . En revanche, cette norme ℓ_1 n'est aucunement un levier pour imposer plus de parcimonie dans la solution.

Pour illustrer cette limite, prenons l'exemple trivial d'un modèle à une seule composante, sans bruit : $\mathbf{y} = s_{n_0}$ et étudions l'estimateur minimisant le critère pénalisé en norme ℓ_1 :

$$\hat{\mathbf{a}}(\mu) = \arg \min_{\mathbf{a} \geq 0} \frac{1}{2} \|\mathbf{s}_{n_0} - \mathbf{S}\mathbf{a}\|^2 + \mu \|\mathbf{a}\|_1, \mu > 0.$$

Soit $\mu^{(0)} = \max_n \mathbf{s}_n^T \mathbf{s}_{n_0} > 0$ (les spectres sont positifs). Pour $\mu \geq \mu^{(0)}$, on a $\hat{\mathbf{a}}(\mu) = \mathbf{0}$. Sous l'hypothèse, réaliste, que ce maximum est atteint en un unique indice n_{\max} , on montre qu'il existe $\mu^{(1)} \in [0, \mu^{(0)}[$ tel que $\forall \mu \in]\mu^{(1)}, \mu^{(0)}[$, $\hat{\mathbf{a}}(\mu)$ possède exactement une composante non nulle¹, celle d'indice n_{\max} , avec

$$n_{\max} = \arg \max_n \mathbf{s}_n^T \mathbf{s}_{n_0} = \arg \max_n \|\mathbf{s}_n\| \cos(\mathbf{s}_{n_0}, \mathbf{s}_n).$$

1. Cet argument est à la base des méthodes homotopiques d'optimisation en norme ℓ_1 , voir notamment [6] pour une version positive.

Les colonnes du dictionnaire n'étant pas normalisées, n_{\max} ne maximise pas seulement la corrélation entre l'atome s_{n_0} et les données s_{n_0} , mais dépend aussi de la norme de s_n , ce qui ne garantit nullement la sélection de $n_{\max} = n_0$. Ainsi, pour le dictionnaire considéré dans cet article (voir les détails en Section 4), pour tout n_0 , les produits $\mathbf{s}_n^T \mathbf{s}_{n_0}$ sont toujours maximaux pour le même atome, celui de norme maximale. Ainsi, la solution en norme ℓ_1 à une composante correspond systématiquement à l'atome de norme maximale, indépendamment de l'atome formant les données !

Notons aussi que ce déficit de la norme ℓ_1 , intrinsèquement lié à la nature non normalisée du dictionnaire, n'est aucunement levé lorsque la contrainte de somme à un est remplacée par une contrainte inégalité $\|\mathbf{a}\|_1 \leq 1$. Pour des raisons similaires, les approches gloutonnes de la parcimonie s'avèrent inefficaces pour prendre en compte la contrainte de somme à un. Ces approches reposant sur l'ajout, à chaque itération, d'une nouvelle composante à une solution initialement nulle, la question d'incorporer efficacement cette contrainte reste, à notre connaissance, ouverte.

3 Reformulation en Programme Mixte en Nombres Entiers

Le problème d'optimisation (2) peut être reformulé exactement sous la forme d'un Programme Mixte en Nombres Entiers (*Mixed Integer Program*, MIP) [7, 4], problèmes d'optimisation mêlant variables continues et discrètes. Par définition, la norme ℓ_0 d'un ensemble des variables est le nombre de ses éléments non nuls. Pour formuler cette norme mathématiquement, on peut utiliser des variables de décision binaires b_n telles que :

$$b_n = 0 \Leftrightarrow a_n = 0. \quad (3)$$

La norme ℓ_0 est alors linéaire en $\mathbf{b} = [b_1, \dots, b_N]^T$: $\|\mathbf{a}\|_0 = \sum_n b_n$. On peut montrer (voir par exemple [4]) que, sous l'hypothèse des contraintes de borne $0 \leq a_n \leq 1$, on a :

$$(3) \Leftrightarrow 0 \leq a_n \leq b_n.$$

Le problème (2) s'écrit alors, de manière exacte :

$$\min_{\mathbf{a} \in [0,1]^N, \mathbf{b} \in \{0,1\}^N} \|\mathbf{y} - \mathbf{S}\mathbf{a}\|^2 \text{ t. q. } \begin{cases} \sum_n b_n \leq K_{\max} \\ 0 \leq \mathbf{a} \leq \mathbf{b} \\ \sum_n a_n = 1 \end{cases} \quad (4)$$

qui est un MIP quadratique à contraintes linéaires. Notons que les problèmes étudiés dans [4] ne sont pas naturellement bornés et requièrent l'introduction d'un paramètre artificiel de borne ($|a_n| < M$, dit « big-M »), dont la valeur impacte fortement le coût de calcul de la solution. À l'inverse, le problème de démixage et ses contraintes naturelles de borne s'avèrent particulièrement adaptés à la reformulation MIP.

L'optimisation mixte en nombres entiers est considérée comme *NP-difficile*, i.e., il n'existe pas d'algorithme polynomial connu qui, étant donné une solution, peut dire si elle est optimale

ou non. Sa résolution a cependant suscité de nombreux travaux dans le domaine de la *recherche opérationnelle*, avec le développement de solveurs spécifiques particulièrement efficaces [8]. Elle repose sur des algorithmes de type *Branch-and-Bound*, qui explorent l'ensemble des solutions réalisables de l'arbre combinatoire d'une manière intelligente et cherchent à éliminer rapidement les solutions les moins pertinentes. Nous utilisons ici le solveur IBM ILOG CPLEX, unanimement reconnu comme l'un des logiciels les plus efficaces sur ces problèmes. Il est important de noter, cependant, que les réglages par défaut de CPLEX se sont avérés inefficaces : en raison de la très forte corrélation entre les spectres du dictionnaire, les solutions obtenues à chaque nœud exploré sont si proches qu'elles provoquent l'arrêt prématuré de l'algorithme. En particulier, la tolérance sur le saut de dualité a dû être abaissée.

4 Résultats de simulations

Nous illustrons maintenant le comportement de l'approche exacte en norme ℓ_0 (ℓ_0 -MIP) sur des problèmes simulés de dé-mélange spectral. Nous utilisons un dictionnaire composé de spectres en réflectance de minéraux susceptibles d'être présents sur des surfaces planétaires [9]. Il s'agit de la base de données du *United States Geological Survey* [10], filtrée de manière à ne garder qu'un seul spectre par classe de minéraux, soit $N = 225$ spectres. L'échantillonnage spectral représente $L = 106$ longueurs d'ondes de 1 à 2.5 microns. Les abondances de ces spectres sont simulées aléatoirement en assurant qu'elles soient positives, de somme égale à un, en contraignant les amplitudes non nulles à dépasser un seuil de 0.04. Leur nombre K_{\max} varie entre 1 et 10, et le rapport signal sur bruit $\text{RSB}_{\text{dB}} = 10 \log \frac{\|\mathbf{S}\mathbf{a}\|_2^2}{\|\boldsymbol{\epsilon}\|_2^2}$ varie entre 60 et 40 dB. Ces valeurs élevées résultent de l'échelle des spectres contenus dans le dictionnaire et de leur positivité et correspondent à des niveaux de bruits réalistes (voir l'exemple de la figure 2). Sur l'ensemble des simulations, pour ℓ_0 -MIP, le paramètre K_{\max} est réglé à sa vraie valeur.

Afin de focaliser sur la capacité des méthodes à détecter les composantes présentes dans les données, nous considérons les mesures d'erreur suivantes, en notant respectivement $\hat{\mathbf{a}}$ et $\hat{\mathbf{a}}$ les abondances estimées et vraies :

- l'erreur quadratique $E_Q = \|\hat{\mathbf{a}} - \hat{\mathbf{a}}\|_2^2$;
- l'erreur d'estimation du support à K composantes $E_{\text{supp}} = \|\text{supp}(\hat{\mathbf{a}}_K) - \text{supp}(\hat{\mathbf{a}})\|_0$, où $\text{supp}(\mathbf{a})_n = 1$ si $a_n \neq 0$ et 0 sinon et, pour les solutions non contraintes en norme ℓ_0 , $\text{supp}(\hat{\mathbf{a}}_K)$ est obtenu en ne conservant que les K plus grandes composantes de $\hat{\mathbf{a}}$.

Nous comparons tout d'abord la solution obtenue par FCLS (critère (1)) et une version similaire « en norme ℓ_1 » où la somme des abondances est contrainte à $\tau \neq 1$. La figure 1 montre ainsi l'erreur E_{supp} en fonction de τ , moyennée sur 300 réalisations des données et sur l'ensemble des valeurs de K_{\max} , et pour plusieurs niveaux de bruit. Comme attendu suite à la discussion de la Section 2, une contrainte en norme ℓ_1 n'améliore pas sensiblement la solution obtenue par FCLS : la valeur τ^* mini-

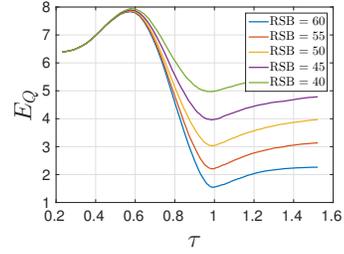


FIGURE 1 – Erreur d'estimation du support pour une approche en norme ℓ_1 ($\|\mathbf{a}\|_1 \leq \tau$) en fonction de τ , pour plusieurs niveaux de bruit.

misant cette erreur passe ainsi de 0.998 lorsque $\text{SNR} = 60$ dB à 0.989 lorsque $\text{SNR} = 40$ dB. Par conséquent, les solutions obtenues sont toujours très proches de celles de FCLS et nous ne présentons donc que celle-ci par la suite.

La figure 2 présente un exemple de résultats obtenus, avec $\text{RSB} = 45$ dB et $K_{\max} = 5$. Sur cet exemple (représentatif

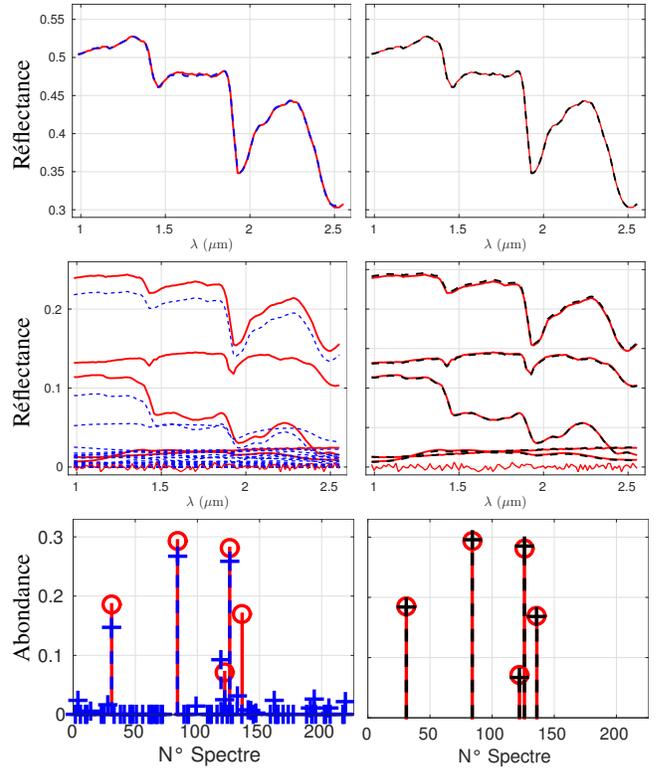


FIGURE 2 – Exemple de résultat ($K_{\max} = 5$ spectres, $\text{RSB} = 45$ dB) : solution de FCLS (gauche) et solution exacte du problème ℓ_0 (droite). En haut : spectre non bruité et sa reconstruction. Au centre, composantes estimées pondérées par les abondances (et bruit sur les données). En bas, abondances estimées. Les traits pleins rouges représentent les vraies valeurs et les pointillés leurs estimations.

des résultats pour des problèmes de cette complexité), ℓ_0 -MIP détecte parfaitement les 5 spectres. En revanche, FCLS ne parvient à détecter correctement que 4 spectres, dont un avec une amplitude estimée très faible, et produit de nombreuses détections erronées. Par conséquent, les abondances des composantes détectées sont sous-estimées.

K_{\max}	1	2	3	4	5	6	7	8	9	10
RSB = 60 dB	0.12	0.19	0.19	0.24	0.28	0.57	1.41	19.91	21.00	66.26 ⁽¹⁾
RSB = 50 dB	0.13	0.26	0.28	0.70	3.56	29.19	33.08	503.1 ⁽⁴⁾	274.5 ⁽⁴⁾	426.4 ⁽¹⁰⁾
RSB = 40 dB	0.16	0.40	0.59	2.13	34.80	612.04 ⁽²⁾	747.4 ⁽⁴⁾	573.2 ⁽¹⁴⁾	582.4 ⁽¹⁴⁾	670.9 ⁽¹⁵⁾

TABLE 1 – Temps de calcul (s) pour l’optimisation du problème reformulé (ℓ_0 -MIP) en fonction du rapport signal sur bruit et de degré de parcimonie, moyennés sur 30 réalisations. Entre parenthèses : nombre de réalisations n’ayant pas fourni la solution optimale en 1000 (s).

La figure 3 présente des résultats quantitatifs sur les erreurs

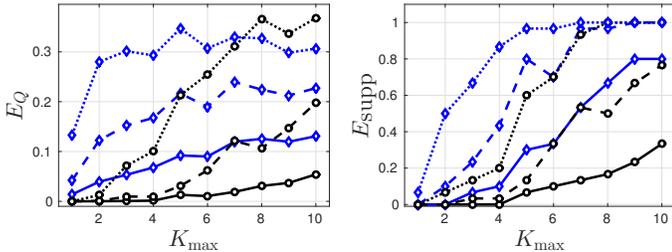


FIGURE 3 – Performances de FCLS (losanges bleus) et de ℓ_0 -MIP (ronds noirs) : erreurs moyennes E_Q et E_{Supp} , sur 30 réalisations des données. RSB = 60 dB (trait plein), 50 dB (tirets), 40 dB (pointillé).

E_Q et E_{Supp} , moyennés sur 30 réalisations des données : ℓ_0 -MIP fournit toujours de meilleurs résultats que FCLS sur les deux mesures jusqu’à $K_{\max} = 8$, et cette différence est d’autant plus marquée que le RSB est élevé. Pour $K_{\max} > 8$ et à RSB faible, les performances se rejoignent. La raison de ce comportement est à la fois d’ordre informationnel (la solution du problème ℓ_0 ne fournit plus la vraie solution, laquelle n’est pas solution du problème en présence de bruit), mais aussi d’ordre computationnel. La Table 1 présente ainsi le temps de calcul moyen de ℓ_0 -MIP, qui augmente fortement avec le nombre de composantes recherchées K_{\max} et le niveau de bruit. En particulier, le solveur trouve et garantit la solution optimale du problème ℓ_0 (2) en moins de 1000 s, temps maximum autorisé, sur toutes les instances jusque $K_{\max} = 5$ pour RSB = 40, $K_{\max} = 7$ pour RSB = 50 et $K_{\max} = 9$ pour RSB = 60.

5 Conclusion

Nous avons montré que la prise en compte de la contrainte de parcimonie en norme ℓ_0 améliore les performances de détection par rapport à la solution classique FCLS, là où l’utilisation de la norme ℓ_1 est peu utile. Même si l’optimisation globale requiert un temps de calcul bien plus élevé que les méthodes sous-optimales, elle reste cependant faisable pour des problèmes dont la complexité est limitée par le nombre de composantes recherchées et le niveau de bruit. C’est également sous cet angle que s’apprécie l’intérêt informationnel de la solution ℓ_0 : sur des problèmes plus complexes, en général, cette dernière ne retrouve pas les vraies abondances. Des simulations supplémentaires ont montré la possibilité d’aborder des problèmes de plus grande dimension, toujours dans ce compromis entre K_{\max} et RSB, grâce notamment aux contraintes de bornes sur les abondances rendant la résolution MIP plus efficace que sur d’autres problèmes parcimonieux.

Nous avons ici supposé la connaissance exacte du nombre de composantes afin de focaliser notre attention sur la capacité de l’approche en norme ℓ_0 à retrouver une solution exactement parcimonieuse. Le réglage automatique de K_{\max} est bien sûr une perspective, une approche consistant à discriminer parmi les solutions obtenues pour différentes valeurs de K_{\max} au moyen d’un critère de sélection de modèle. De même, le comportement de cette approche exacte en présence d’une solution non exactement parcimonieuse serait d’un grand intérêt pratique. Enfin, la souplesse de la formulation MIP, par l’introduction de variables de décision, permet également d’envisager la prise en compte de contraintes habituellement abordées de manière approchée, de type parcimonie structurée ou encore de valeurs minimales imposées aux abondances non nulles.

Remerciements. Ce travail a été partiellement financé par l’Agence Nationale de la Recherche, projet MIMOSA ANR-16-CE33-0005 et le CNRS *via* le défi Imag’in MultiPlanNet.

Références

- [1] D.C. Heinz and Chein-I-Chang, “Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 3, Mar 2001.
- [2] M.D. Iordache, J.M. Bioucas-Dias, and A Plaza, “Sparse unmixing of hyperspectral data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, June 2011.
- [3] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, “Hyperspectral unmixing overview : Geometrical, statistical, and sparse regression-based approaches,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, April 2012.
- [4] S. Bourguignon, J. Ninin, H. Carfantan, and M. Mongeau, “Exact sparse approximation problems via mixed-integer programming : Formulations and computational performance,” *IEEE Trans. Signal Process.*, vol. 64, no. 6, March 2016.
- [5] R.B. Singer and T.B. McCord, “Mars - Large scale mixing of bright and dark surface materials and implications for analysis of spectral reflectance,” in *Lunar and Planetary Science Conference Proceedings*, 1979, vol. 10.
- [6] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, “Least angle regression,” *Annals of Statistics*, vol. 32, 2004.
- [7] S. Jögar and M. Pfetsch, “Exact and approximate sparse solutions of underdetermined linear equations,” *SIAM J. Sci. Comp.*, vol. 31, no. 1, 2008.
- [8] R.E. Bixby, “A brief history of linear and mixed-integer programming computation,” *Doc. Math.*, vol. Optimization Stories, 2012.
- [9] F. Schmidt, M. Legendre, and S. Le Mouëlic, “Minerals detection for hyperspectral images using adapted linear unmixing : LinMin,” *Icarus*, vol. 237, 2014.
- [10] R.N. Clark, G.A. Swayze, R. Wise, K.E. Livo, T. Hoefen, R.F. Kokaly, and S.J. Sutley, “USGS digital spectral library splib05a,” *US Geological Survey, Digital Data Series*, vol. 231, 2003.